

Molecular Evolution of the Human Enteroviruses: Correlation of Serotype with VP1 Sequence and Application to Picornavirus Classification

M. STEVEN OBERSTE,* KAIJA MAHER, DAVID R. KILPATRICK, AND MARK A. PALLANSCH

Respiratory and Enteric Viruses Branch, Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia 30333

Received 3 September 1998/Accepted 30 November 1998

Sixty-six human enterovirus serotypes have been identified by serum neutralization, but the molecular determinants of the serotypes are unknown. Since the picornavirus VP1 protein contains a number of neutralization domains, we hypothesized that the VP1 sequence should correspond with neutralization (serotype) and, hence, with phylogenetic lineage. To test this hypothesis and to analyze the phylogenetic relationships among the human enteroviruses, we determined the complete VP1 sequences of the prototype strains of 47 human enterovirus serotypes and 10 antigenic variants. Our sequences, together with those available from GenBank, comprise a database of complete VP1 sequences for all 66 human enterovirus serotypes plus additional strains of seven serotypes. Phylogenetic trees constructed from complete VP1 sequences produced the same four major clusters as published trees based on partial VP2 sequences; in contrast to the VP2 trees, however, in the VP1 trees strains of the same serotype were always monophyletic. In pairwise comparisons of complete VP1 sequences, enteroviruses of the same serotype were clearly distinguished from those of heterologous serotypes, and the limits of intraserotypic divergence appeared to be about 25% nucleotide sequence difference or 12% amino acid sequence difference. Pairwise comparisons suggested that coxsackie A11 and A15 viruses should be classified as strains of the same serotype, as should coxsackie A13 and A18 viruses. Pairwise identity scores also distinguished between enteroviruses of different clusters and enteroviruses from picornaviruses of different genera. The data suggest that VP1 sequence comparisons may be valuable in enterovirus typing and in picornavirus taxonomy by assisting in the genus assignment of unclassified picornaviruses.

Human enteroviruses (family *Picornaviridae*) infect millions of people worldwide each year, resulting in a wide range of clinical outcomes ranging from inapparent infection to mild respiratory illness (common cold), hand-foot-and-mouth disease, acute hemorrhagic conjunctivitis, aseptic meningitis, myocarditis, severe neonatal sepsis-like disease, and acute flaccid paralysis (reviewed in references 43 and 45). In the United States, enteroviruses are responsible for 30,000 to 50,000 meningitis hospitalizations per year as a result of 30 million to 50 million infections. Serologic studies have distinguished 66 human enterovirus serotypes on the basis of an antibody neutralization test (43), and additional antigenic variants have been defined within several of the serotypes on the basis of reduced or nonreciprocal cross-neutralization between prototype and variant strains (6, 8, 68, 71, 72). On the basis of their pathogenesis in humans and experimental animals, the enteroviruses were originally classified into four groups, polioviruses, coxsackie A viruses (CA), coxsackie B viruses (CB), and echoviruses, but it was quickly realized that there were significant overlaps in the biological properties of viruses in the different groups (8). The more recently isolated enteroviruses have been named with a system of consecutive numbers: EV68, EV69, EV70, and EV71 (42).

A comparison of nucleotide and deduced amino acid sequences at the 5' end of VP2 has identified four major phylogenetic groups within the *Enterovirus* genus: CA16-like viruses (cluster A), a CB-like group containing all CB and echoviruses

as well as CA9 and EV69 (cluster B), poliovirus-like viruses (cluster C), and EV68 and EV70 (cluster D) (23, 24, 49, 53, 54, 73). However, pairwise alignments and phylogenetic analyses within these groups demonstrated that the VP2 sequence does not fully correlate with serotype, as viruses known to belong to the same serotype often failed to cluster together (2, 49). (E22 and E23 are genetically distinct from enteroviruses [24], and their reclassification into a separate genus has been proposed [45]).

VP1 is the most external and immunodominant of the picornavirus capsid proteins (58). A number of major neutralization sites reside in the VP1 proteins of many picornaviruses (reviewed in references 40 and 44), but the specific epitopes responsible for serotype specificity and intratypic variation have not been identified. Similarly, the genetic correlates of serotype identity remain unknown. If the important serotype-specific neutralization sites reside in VP1, then the VP1 sequence or some portion thereof would be predicted to correlate with serotype. Studies on the three serotypes of poliovirus have shown that a partial VP1 sequence correlates well with serotype (32). In addition, genetic lineages based on the VP1 sequence can be used to define poliovirus reservoirs and chains of transmission (reviewed in reference 30). To test whether the VP1 sequence might be applied to the classification of nonpolio enteroviruses and to the analysis of the phylogenetic relationships among the human enteroviruses, we determined the complete VP1 nucleotide sequences for 47 human enterovirus prototypes and 10 well-characterized antigenic variants. These data, together with previously available sequences, comprise a database of complete VP1 sequences for all known human enterovirus serotypes and 12 natural antigenic variants. This database will be useful for molecular epidemiologic studies of

* Corresponding author. Mailing address: Centers for Disease Control and Prevention, 1600 Clifton Rd. NE, Mailstop G-17, Atlanta, GA 30333. Phone: (404) 639-2751. Fax: (404) 639-4011. E-mail: mbo2@cdc.gov.

TABLE 1. Enterovirus VP1 sequences used in sequence comparisons and phylogenetic analyses

Serotype ^a	Strain ^b	GenBank accession no. (sequence reference) ^c
CA1	Tompkins	AF081293
CA2	Fleetwood	L28146 (54)
CA3	Olson	AF081294
CA4	High Point	AF081295
CA5	Swartz	AF081296
CA6	Gdula	AF081297
CA7	AB-IV	AF081298
CA8	Donovan	AF081299
CA9	Griggs	D00627 (5)
CA10	Kowalik	AF081300
CA11	Belgium-1	AF081301
CA12	Texas-12	AF081302
CA13	Flores	AF081303
CA14	G-14	AF081304
CA15	G-9	AF081305
CA16	G-10	U05876 (52)
CA17	G-12	AF081306
CA18	G-13	AF081307
CA19	8663	AF081308
CA20	IH-35	AF081309
CA21	Kuykendall	D00538 (22)
CA22	Chulman	AF081310
CA24	Joseph	AF081311
CA24v	EH24/70	D90457 (65)
CB1	Conn-5	M16560 (26)
CB2	Ohio-1	AF081312
CB3	Nancy	M16572 (38)
CB4	JVB	D00149 (28)
CB5	Faulkner	X67706 (74)
CB6	Schmitt	AF081313
E1	Farouk	AF081314
E2	Cornelis	AF081315
E3	Morrissey	AF081316
E4	Pesacek	AF081317
E4	Du Toit	AF081318
E4	Shropshire	AF081319
E5	Noyce	AF081320
E6	Charles	U16283 (20)
E6	D'Amori	AF081321
E6'	Cox	AF081322
E6''	Burgess	AF081323
E7	Wallace	AF081324
E8 ^d	Bryson	AF081325
E9	Hill	X84981 (75)
E9	Barty	X92886 (76)
E11	Gregory	X80059 (11)
E11'	Silva	AF081326
E12	Travis	X79047 (36)
E13	Del Carmen	AF081327
E14	Tow	AF081328
E15	CH96-51	AF081329
E16	Harrington	X89545 (23)
E17	CHHE-29	AF081330
E18	Metcalf	AF081331
E19	Burke	AF081332
E20	JV-1	AF081333
E21	Farina	AF081334
E22	Harris	S45208 (23)
E23	Williamson	AF055846 (50)
E24	DeCamp	AF081335
E25	JV-4	AF081336
E26	Coronel	AF081337
E27	Bacon	AF081338
E29	JV-10	AF081339
E30	Bastianni	AF081340

Continued

TABLE 1—Continued

Serotype ^a	Strain ^b	GenBank accession no. (sequence reference) ^c
E30	Frater	AF081341
E30	Giles	AF081342
E30	PR-17	AF081343
E31	Caldwell	AF081344
E32	PR-10	AF081345
E33	Toluca-3	AF081346
E34 ^e	DN-19	AF081347
EV68	Fermon	AF081348
EV69	Toluca-1	AF081349
EV70	J670/71	D00820 (59)
EV71	BrCr	U22521 (3)
PV1	Mahoney	J02281 (34, 55)
PV1	Sabin	V01150 (14, 16, 47)
PV2	Lansing	M12197 (37)
PV2	Sabin	X00595 (66)
PV3	Leon	K01392 (62)
PV3	Sabin	X00596 (66)
BEV1	VG-5-27	D00214 (15)
BEV2a	RM-2	X79369 (41)
BEV2b	PS-87	X79368 (41)
PEV9	UKG/410/73	Y14459 (51)
SVDV	H/3'76	D00435 (28)
HRV2	Unknown	X02316 (61)
HRV14	Unknown	K02121, X01087 (4, 63)
HAV	HM-175	M14707 (7)

^a CA, coxsackie A virus; CB, coxsackie B virus; E, echovirus; EV, enterovirus; PV, poliovirus; BEV, bovine enterovirus; PEV, porcine enterovirus; SVDV, swine vesicular disease virus; HRV, human rhinovirus; HAV, human hepatitis A virus.

^b Variant strains of recognized serotypes and attenuated poliovirus vaccine strains are also given.

^c Original sequence reference(s) if not sequenced in this work.

^d Reclassified as a strain of E1 (9, 20).

^e Reclassified as a strain of CA24 (60).

enteroviral disease outbreaks, to obtain a better understanding of the genetic correlates of serotype, and for the development of enteroviral molecular diagnostic reagents.

MATERIALS AND METHODS

Viruses, RNA extraction, and reverse transcription-PCR. The viruses used for sequence analysis and phylogenetic reconstruction are listed in Table 1. RNA isolation and reverse transcription-PCR were carried out as described previously (48). Briefly, viral RNA was extracted from infected cell culture supernatant or 10% infected mouse brain homogenate with Trizol LS (Life Technologies, Inc., Gaithersburg, Md.), and cDNA was synthesized by use of a random hexamer primer and a SuperScript preamplification kit (Life Technologies, Inc.). From each viral cDNA, an amplicon of approximately 900 to 950 bp, encompassing the 3' end of VP3, all of VP1, and the 5' end of 2A, was amplified by PCR with primers for VP3 and 2A (Table 2). For some viruses, VP1 was amplified as two overlapping fragments with internal VP1 primers as well as the VP3 and 2A primers. The PCR products were gel isolated and purified for sequencing with a QIAquick gel extraction kit (Qiagen, Inc., Santa Clarita, Calif.) and sequenced on an automated DNA sequencer with fluorescent dideoxy chain terminators (PE-Applied Biosystems, Foster City, Calif.). Complete VP1 PCR products of viruses for which VP1 primers were not available were cloned into pGEM-T (Promega Corp., Madison, Wis.), and nested-deletion subclones were constructed with an Erase-a-Base kit (Promega). For each virus, at least two independent clones were sequenced by automated methods as described above.

Sequence analysis. Pairwise nucleotide and amino acid sequence identities were calculated by alignment of all possible sequence pairs with the program Gap (Wisconsin Sequence Analysis Package, version 9.1; Genetics Computer Group, Inc., Madison, Wis.). For phylogenetic reconstructions, groups of nucleotide sequences were aligned with Pileup (Genetics Computer Group). The alignment was manually adjusted to account for codon boundaries, optimal alignment among closely related sequences, and optimal alignment of highly conserved amino acid motifs. Phylogenetic relationships were inferred with the programs Neighbor and DNAPars (PHYLIP version 3.57 [17]) and Puzzle (ver-

TABLE 2. Primers used for PCR amplification of the VP1 region of enteroviruses

Primer	Sequence ^a	Gene	Position ^b
008	GCRTGCAATGAYTTCTCWGT	VP3	2411–2430
009	NGCNCCDATTGNTGSCC	2A	3409–3391
011	GCICIGAYTGITGICCAA	2A	3408–3389
012	ATGTAYGTICCCICGIGG	VP1	2951–2970
013	GGIGCRTTICCYTCIGTCCA	VP1	3051–3032
019	ACRTGICIGTYTGCATIGT	VP1	2676–2657
035	AWITTYTAYGAYGGITGG	VP1	3098–3115
036	TAIAIGTICCATRTRTT	VP1	3201–3182
040	ATGTAYRTICCMICGIGC	VP1	2951–2970
041	GGIGGIGGRTICGTIAKYTT	VP1	3054–3035
045	GAIGARAAAYCTIATIGARAC	VP1	2648–2667
046	CCCATIAKRTCIATRTCCC	VP1	2820–2801
050	GTRCTYACIAIAGRTCYCT	2A	3513–3494
051	TSAARYTGTGCAARGACAC	VP3	2429–2448
052	STGYCCAGATTTCAGTGT	VP3	2413–2430
053	GGNACNCAYRTNATHTGGGA	VP3	2216–2235
054	GCCITRTTITGRTGICCAA	2A	3408–3389
055	GGIACICAYRTIRTITGGGA	VP3	2216–2235

^a Sequences are shown 5' to 3' with standard codes.

^b Nucleotide sequence coordinates are given relative to the published sequence of PV1-Mahoney (34, 55) for orientation only. For most of the primers, specificity is limited to a subset of the human enteroviruses; not all virus-primer pair combinations have been tested.

sion 4.0 [64]). The maximum-likelihood method of Kishino and Hasegawa (33), with a transition/transversion (Ts/Tv) ratio of 8.0, was used to construct a distance matrix for neighbor-joining analysis. The statistical significance of phylogenies constructed with Neighbor and DNAPars was estimated by bootstrap analysis with 100 pseudoreplicate data sets. Puzzle was executed by use of the distance method of Kishino and Hasegawa (33), with a Ts/Tv ratio of 8.0, and the reliability of phylogenetic reconstructions was estimated by use of 1,000 puzzling steps. Branch lengths of the neighbor-joining trees were calculated by the maximum-likelihood method with Puzzle. The robustness of subtrees within cluster B (CB-like group) was confirmed by construction of a Puzzle tree containing each of the subtree taxa, as well as their nearest sibling taxa, for a total of 14 sequences. The sibling taxa were those with the highest amino acid identity scores compared with the taxa in the subtree of interest. For subtree analysis, Puzzle was executed with Ts/Tv ratios of 1.0, 6.0, and 10.0, and well-supported Puzzle trees were taken as evidence in support of the subtree in the original phylogram.

Nucleotide sequence accession numbers. The sequences reported here were deposited in the GenBank sequence database under accession no. AF081293 to AF081349.

RESULTS

Complete VP1 nucleotide sequences were determined for 57 human enterovirus strains for which VP1 sequences were not previously available (Table 1). Forty-seven of the strains were prototype strains for recognized human enterovirus serotypes (43). The other 10 sequenced strains were well-characterized antigenic variants which, while antigenically distinct from their respective prototype strains, were similar enough to the prototype strains to have been considered of the same serotype (8, 43). Combined with the 21 previously available complete enterovirus VP1 sequences, the 57 sequences reported here comprise the first collection of complete gene sequences representing each of the 66 human enterovirus serotypes.

The boundaries of the newly sequenced VP1 genes were predicted by comparison of the nucleotide and deduced amino acid sequences with those of previously characterized enteroviruses. Human enterovirus VP1 sequences varied in length from 834 to 951 nucleotides (278 to 317 amino acids). The CB had the shortest predicted VP1 amino acid sequences (278 to 298 amino acids), while EV68 and EV70 had the longest ones (312 and 317 amino acids). The newly determined enterovirus sequences were compared with previously available human

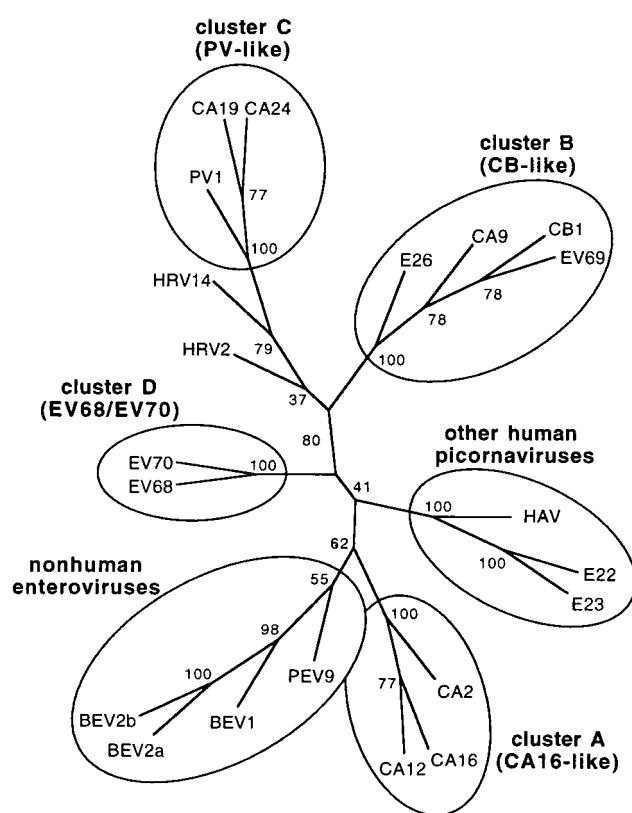


FIG. 1. Consensus phylogenetic tree for representative human picornaviruses and nonhuman enteroviruses constructed by the neighbor-joining method (17) with a Ts/Tv ratio of 8.0. Numbers at nodes represent the percentage of 100 bootstrap pseudoreplicates that contained the cluster distal to the node. Major clusters supported by bootstrap values of at least 67% are enclosed by circles. For clarity, branch lengths are not drawn to scale. PV, poliovirus.

enterovirus VP1 sequences and with the sequences of other closely related picornaviruses, including E22 and E23, provisionally reclassified as the only members of the genus *Parechovirus* (45); porcine enterovirus 9; bovine enterovirus types 1 and 2; human rhinovirus types 2 and 14; and human hepatitis A virus.

To assess the broad relationships among the enteroviruses and other human picornaviruses, a phylogenetic tree was constructed with representatives of each of the four human enterovirus clusters (A: CA2, CA12, and CA16; B: CA9, CB1, E26, and EV69; C: PV1, CA19, and CA24; D: EV68 and EV70), the nonhuman enteroviruses (BEV1, BEV2a, BEV2b, and PEV9), and other human picornaviruses (E22, E23, HAV, HRV2, and HRV14). As expected, the human enteroviruses clustered into four major groups (Fig. 1), consistent with published enterovirus phylogenies (23, 49, 53, 54, 73). Picornaviruses from the genera *Parechovirus* and *Hepatovirus* were distinct from the enterovirus clusters. The human rhinoviruses HRV2 and HRV14 clustered among the human enteroviruses, consistent with previous phylogenies, but the precise position of HRV2 was poorly supported (bootstrap value, 37%). The nonhuman enteroviruses, PEV9, BEV1, BEV2a, and BEV2b, formed a monophyletic group distinct from but related to cluster A. The human enterovirus clusters were very strongly supported, with bootstrap values of 100%, and the relationship between clusters B and C was also well supported (80%). The relationship of the nonhuman enteroviruses to cluster A was

supported by a bootstrap value of 62%. In some cases, PEV9 fell outside this group, resulting in the low bootstrap value.

To determine the phylogenetic relationships among individual prototype viruses and, where available, prototypes and their antigenic variants, intracluster phylogenetic trees were constructed with several phylogeny reconstruction algorithms (neighbor-joining, maximum parsimony, and maximum likelihood) and included one sequence from each of the heterologous human enterovirus clusters as an outgroup. Within each of the four major clusters, trees constructed by different methods were congruent in overall structure and in general clustering patterns but often differed slightly in the order of one or more distal branches or in the bootstrap support for specific nodes. Viruses in cluster A segregated into three distinct subgroups—(i) CA7, CA14, CA16, and EV71; (ii) CA3, CA4, CA6, CA8, and CA10; and (iii) CA5 and CA12—with 59 to 99% bootstrap support (Fig. 2A). CA2 appeared to be distinct from the other viruses in cluster A. Cluster C viruses segregated into four subgroups—(i) CA1, CA19, and CA22; (ii) CA21, CA24, CA24v, and E34; (iii) CA11 and CA15; and (iv) CA13, CA17, CA18, CA20, PV1, PV2, and PV3—with bootstrap support of 67 to 100% (Fig. 2C). Within cluster B, some of the viruses clustered into subgroups, but few of the subgroups were well supported by bootstrap analysis (Fig. 2B). Stable subgroups included (i) E3 and E12; (ii) E11, E11', and E19; (iii) E2 and E15; (iv) E13 and EV69; (v) the six CB serotypes; (vi) E1, E8, E4-Pesacek, E4-DuToit, and E4-Shropshire; (vii) E6-D'Amori, E6-Charles, E6'-Cox, and E6"-Burgess; and (viii) E21, E25, E29, E30-Bastianni, E30-Frater, E30-Giles, and E30-PR-17. The other viruses could not be reliably subgrouped, as the bootstrap values were extremely low. For every cluster, all serotypes which were represented by more than one isolate were monophyletic. In most cases, bootstrap support was strong (60 to 98%), but the value for CA24 strains was only 43%.

Sequence relationships within a serotype, within a cluster, between clusters, and between human enteroviruses and other picornaviruses were analyzed by comparison of the nucleotide and deduced amino acid sequences of all possible sequence pairs. The relationships were visualized by plotting the frequency of pairwise identity scores versus percent identity, rounded down to the nearest integer, as a histogram (Fig. 3). For both the nucleotide (Fig. 3A) and amino acid (Fig. 3B) pairwise identity distributions, the scores fell into four categories. The highest scores (nucleotide identity, $\geq 75\%$; amino acid identity $\geq 88\%$) depicted relationships among viruses of the same serotype (e.g., the four E30 strains) or among prototype viruses that have been proposed to be homologous based on antigenic relatedness (e.g., CA13 and CA18). Nucleotide identity scores for pairwise comparisons within a major cluster ranged from 48.9 to 73.2% and defined a peak that was clearly delineated from that of the homologous pairs and from the peak of scores comparing viruses of different phylogenetic clusters (Fig. 3A). Cluster A scores ranged from 58.5 to 73.2%, while cluster C scores ranged from 55.9 to 70.6%. Viruses in cluster B appeared to be somewhat more heterogeneous, with scores ranging from 48.9 to 71.8%. Scores for the heterologous comparison peak ranged from 42.1 to 64.5% nucleotide identity. The final peak, containing the lowest scores, represented comparisons of viruses of different genera within the family *Picornaviridae*. In the amino acid identity distribution (Fig. 3B), the heterologous cluster peak appeared to be composed of two overlapping peaks. The peak with higher scores represented comparisons of viruses from phylogenetically related clusters (e.g., clusters B and C), whereas the peak with lower

scores represented comparisons of viruses from more distant clusters (e.g., clusters A and B).

DISCUSSION

VP1 is the major surface-accessible protein in the mature picornavirus virion; it is arrayed around the fivefold axis of symmetry of the icosahedral virion (1, 21, 39, 58). VP2 and VP3 comprise the remainder of the virion surface. Each of the capsid proteins is composed of conserved elements that form the β -barrel structural elements of the capsid, with variable loops between the β -barrel structures (reviewed in reference 44). Many of the loops are exposed on the virion surface, and studies of monoclonal antibody-resistant mutants have shown that a number of the loops contribute to specific antigenic neutralization sites. VP1 contributes to all three of the major neutralization sites that have been identified on the poliovirus surface, whereas VP2 and VP3 contribute to two and one of the sites, respectively. For example, the B-C loop is one of five VP1 loops forming VP1 antigenic site 1. Replacement of the VP1 B-C loop of CB3 by that of CB4 through site-directed mutagenesis produced a viable virus with a mixed neutralization phenotype, demonstrating the presence of a serotype-specific antigenic neutralization site in the B-C loop (57). Our sequencing studies confirmed the presence of sequence domains that are conserved among all members of the *Enterovirus* genus, as well as intervening domains that vary in sequence between strains of different serotypes and in some cases within a serotype. Due to the complexity of the three-dimensional structure of the enterovirus capsid and the fact that most of the neutralization sites are discontinuous, it is not possible to correlate specific VP1 residues with antigenic sites responsible for serotype specificity. Sequence comparisons and phylogenetic reconstructions suggest that VP1 contains serotype-specific information that can be used for virus identification and evolutionary studies. For the polioviruses, serotype-specific sequences in VP1 have been exploited to produce type-specific molecular diagnostic reagents (see below).

Pairwise sequence identity has been applied to the taxonomy of plant viruses in the family *Potyviridae*, with clear distinction of viruses of the same strain, viruses of different strains within a genus, and viruses of different genera (67). Our data suggest that a similar system based on the VP1 sequence can be used to classify viruses within the family *Picornaviridae* (Fig. 3). Enteroviruses of the same serotype were clearly distinguished from those of heterologous serotypes, and the limits of intra-serotypic divergence appeared to be about 25% nucleotide sequence difference or 12% amino acid sequence difference. Likewise, strains of a homologous serotype clustered together in phylogenetic analyses (Fig. 2). Sequencing of additional strains within several different serotypes will provide more extensive data to determine whether this distinction is valid and will provide a more accurate measure of the serotype boundaries. Since enterovirus serotypic differentiation is based on neutralization and the VP1 sequence correlates with neutralization type, it is logical to assume that molecular diagnostics targeted to the VP1 coding region should give typing results that also correlate with the serotype determined by neutralization with type-specific antisera. The molecular distinction between serotypes has obvious applications to the typing of enterovirus isolates in the clinical laboratory. Molecular assays directed to specific sequences in VP1 have already been applied to the serotyping, genotyping, and group identification of polioviruses (12, 13, 30–32, 69, 70). Pairwise identity scores have also placed viruses of the same cluster in a single frequency peak, demonstrating that gross intercluster sequence

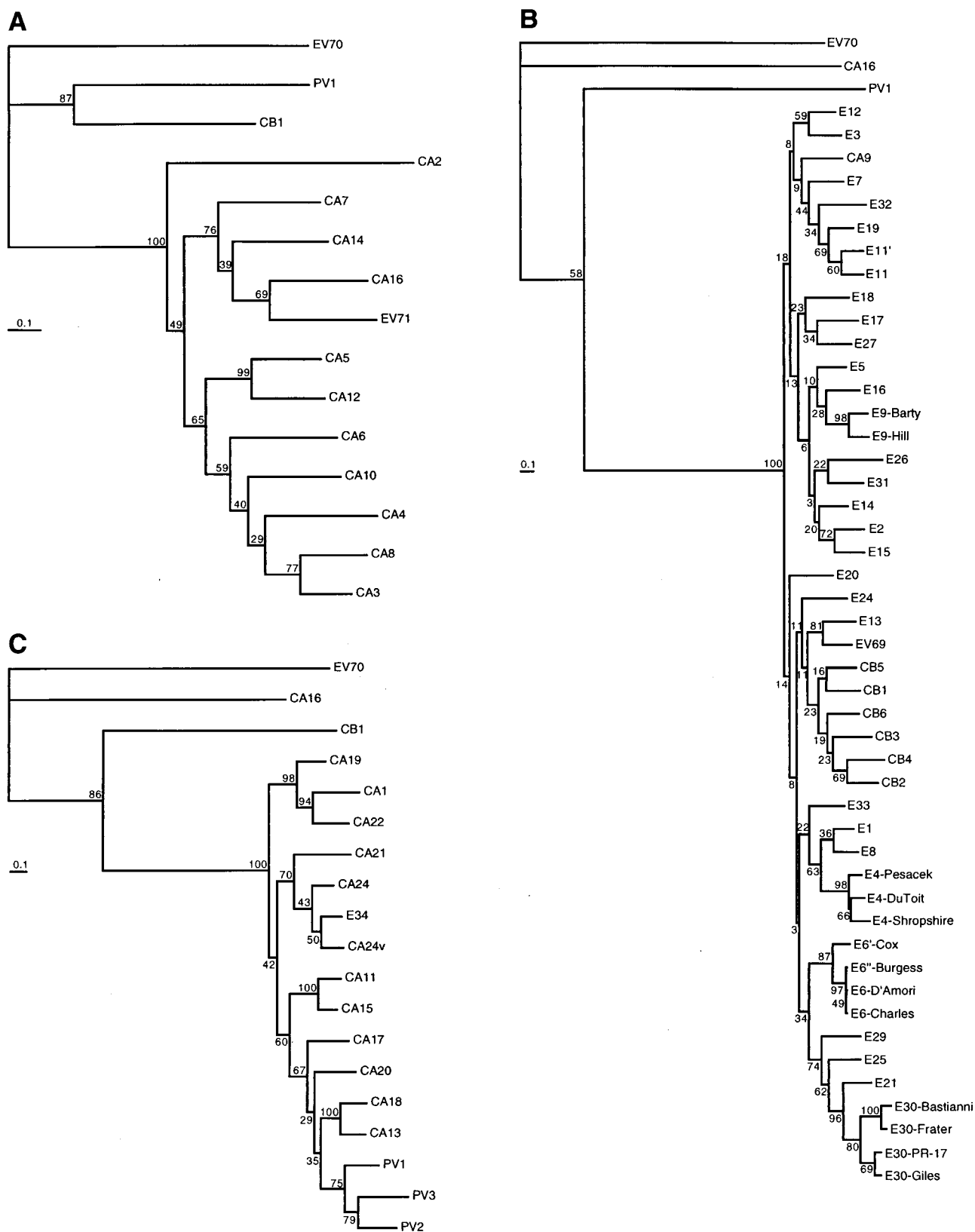


FIG. 2. Consensus phylogenetic trees for human enterovirus clusters A, B, and C, constructed by the neighbor-joining method (17) with a Ts/Tv ratio of 8.0. Branch lengths are proportional to genetic distances calculated by the maximum-likelihood method implemented in Puzzle (64). Numbers at nodes represent the percentage of 100 bootstrap pseudoreplicates that contained the cluster distal to the node. Genetic distance is indicated by the bar at lower left in each panel. (A) CA16-like viruses (cluster A). (B) CB-like viruses (cluster B). (C) Poliovirus-like viruses (cluster C).

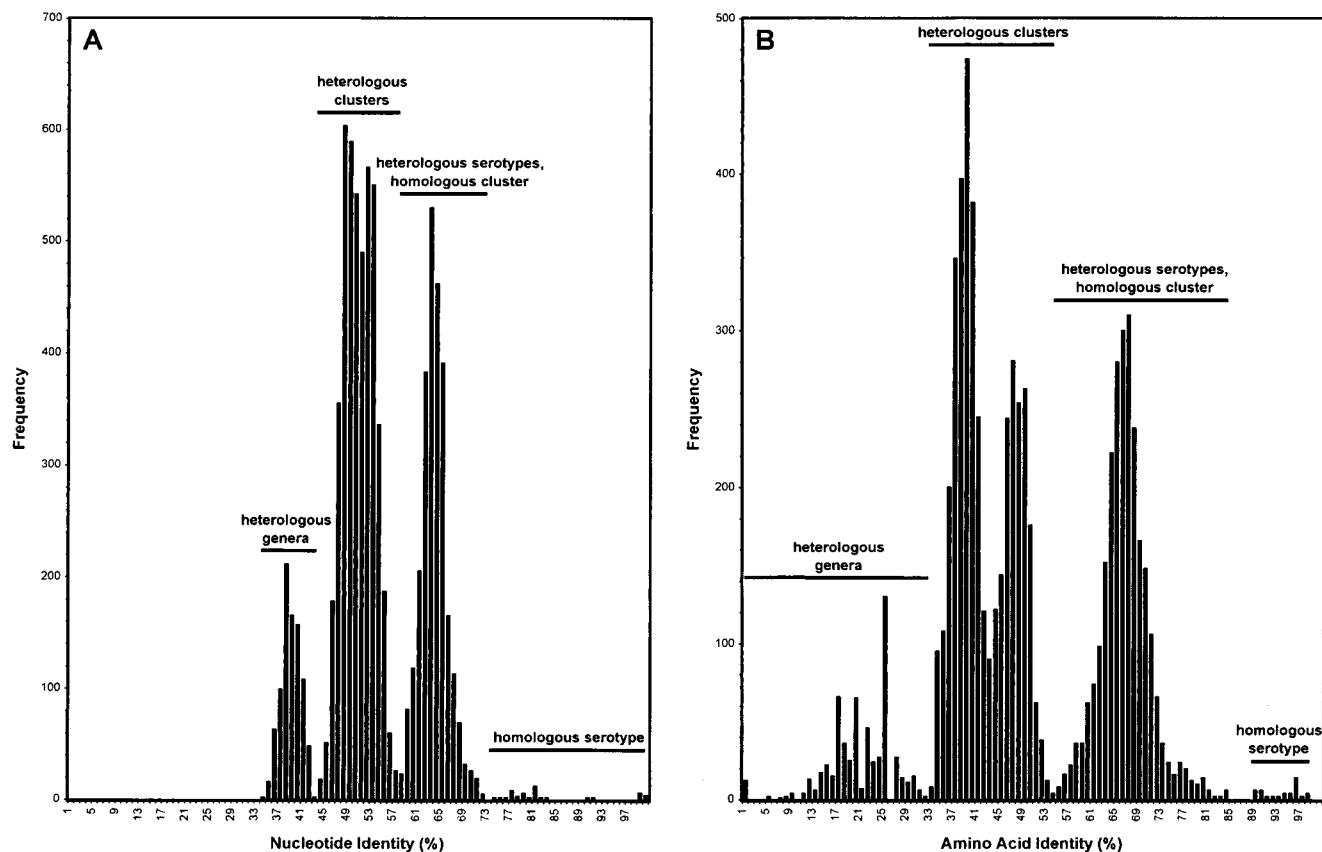


FIG. 3. Frequency distribution of pairwise identity scores for comparison of VP1 nucleotide (A) and deduced amino acid (B) sequences.

differences exist, whereas differences within a cluster occur on a smaller scale. The molecular distinction between clusters suggests that it may be useful to extend the taxonomic classification of enteroviruses to include each of the clusters as a subgenus. Similarly, VP1 sequence comparisons may prove valuable in the assignment of unclassified picornaviruses to one of the existing genera or point out the need for the introduction of new genera.

The CB were originally distinguished from the CA and polioviruses on the basis of differences in pathogenesis when inoculated intracerebrally into newborn mice. The echoviruses were, by definition, apathogenic in newborn mice; however, there are examples of echovirus isolates that are mouse virulent. In phylogenetic trees based on the partial sequence of VP2 (or VP4-VP2), the CB failed to cluster together, being interspersed among the echoviruses in a large CB-echovirus group (23, 49). In addition, analysis of 5' nontranslated region sequence of seven CB5 clinical isolates showed that there is little or no genetic linkage between the 5' nontranslated region sequences and serotype, probably due to a high frequency of recombination, whereas VP1-VP2A junction sequences are monophyletic and are correlated with serotype (35). In the present work, VP1 trees contained a monophyletic group consisting only of CB, suggesting that biological properties exclusive to the CB may be partially or completely encoded within the VP1 gene. Although bootstrap support for this subgroup was relatively low in the overall cluster B tree (Fig. 2B), additional tree reconstruction with only sequences most closely related to the CB supported the existence of a distinct CB subgroup within cluster B (data not shown). The nature of the

CB-specific determinants remains unknown but could include receptor specificity and, hence, host range, cell or tissue tropism, and pathogenesis, as structural studies have shown that VP1 forms part of the picornavirus receptor-binding pocket (58).

CA1, CA19, and CA22 are the only enterovirus serotypes that have never been successfully adapted to growth in cell cultures, instead requiring passage by intracranial inoculation of suckling mice (18, 29, 43). In parsimony and neighbor-joining analyses with partial VP2 sequences, these three viruses formed a monophyletic cluster within the poliovirus-like group, but with low bootstrap support (49). The maximum-likelihood method produced a VP2 tree with a star topology within the poliovirus-like cluster and failed to show specific clustering of CA1, CA19, and CA22. When complete VP1 sequences were used, the same algorithms consistently produced trees in which CA1, CA19, and CA22 formed a well-resolved monophyletic cluster within the poliovirus-like group, with bootstrap support of 77 to 100%, suggesting that VP1 may play a role in the common host range and/or cell tropism of these three viruses.

E8 (strain Bryson) was isolated in 1953 and initially classified as a separate enterovirus serotype (8, 56). Later serologic studies demonstrated the antigenic relatedness between E8 and E1 (9, 20), and E8 has been designated a strain of the E1 serotype (45). Our genetic comparison of E1 and E8 agrees with the previous antigenic studies, with E1 and E8 being 76.0% identical in nucleotide sequence and 93.2% identical in amino acid sequence, and supports the reclassification of E8 as a variant of E1, based on the homologous serotype boundary described

above. The VP1 nucleotide sequences of CA11 (Belgium-1) and CA15 (G-9) were 80.1% identical to one another (96.7% amino acid identity). Partial VP2 sequences had shown that the two viruses were 100% identical in the 50 amino acids at the amino terminus of VP2 (49). Similarly, the complete VP1 nucleotide sequences of CA13 (Flores) and CA18 (G-13) were 77.2% identical to one another (95.1% amino acid identity). Antigenic cross-reactivity has been noted between CA11 and CA15 and between CA13 and CA18 (9). Taken together, the serologic and genetic relationships suggest that CA15 may be a strain of CA11 and that CA18 may be a strain of CA13. In contrast, antigenic cross-reactivity has been reported for CA3 and CA8 (10), but their VP1 amino acid sequences were only 83.3% identical, suggesting that they are genuinely distinct serotypes that share a common epitope(s).

We have demonstrated the general utility of the enterovirus VP1 sequence database by its application to basic problems in enterovirus classification, phylogeny, and evolution. Based on the success of poliovirus molecular diagnostics targeted to VP1 (12, 13, 31, 32, 69, 70) and the shortcomings of current molecular methods for identifying nonpoliovirus enteroviruses (2), we conclude that future enterovirus molecular diagnostic development efforts should be targeted to the genomic region encoding VP1. Hybridization, PCR, and sequencing assays targeted to other regions of the genome may still be useful in identifying and characterizing intertypic enterovirus recombinants. The existence of a complete VP1 sequence database that includes representatives of all human enterovirus serotypes will facilitate the development of generic and serotype-specific molecular reagents for the rapid diagnosis of enterovirus infections, for the detailed molecular epidemiologic study of enterovirus disease outbreaks, for the characterization of newly discovered enteroviruses, and for the study of enterovirus evolution within and among the 66 human enterovirus serotypes.

REFERENCES

1. Archarya, R., E. Fry, D. Stuart, G. Fox, D. J. Rowlands, and F. Brown. 1989. The three dimensional structure of foot and mouth disease virus at 2.9 Å resolution. *Nature* **338**:709–716.
2. Arola, A., J. Santti, O. Ruuskanen, P. Halonen, and T. Hyypiä. 1996. Identification of enteroviruses in clinical specimens by competitive PCR followed by genetic typing using sequence analysis. *J. Clin. Microbiol.* **34**:313–318.
3. Brown, B. A., and M. A. Pallansch. 1995. Complete nucleotide sequence of enterovirus 71 is distinct from poliovirus. *Virus Res.* **39**:195–205.
4. Callahan, P. L., S. Mizutani, and R. J. Colonna. 1985. Molecular cloning and complete sequence determination of RNA genome of human rhinovirus type 14. *Proc. Natl. Acad. Sci. USA* **82**:732–736.
5. Chang, K., P. Auvinen, T. Hyypiä, and G. Stanway. 1989. The nucleotide sequence of coxsackievirus A9: implications for receptor binding and enterovirus classification. *J. Gen. Virol.* **70**:3269–3280.
6. Cherry, J. D., A. M. Lerner, J. O. Klein, and M. Finland. 1963. Echovirus 11 virus infections associated with exanthems. *Pediatrics* **32**:509–516.
7. Cohen, J. I., J. R. Ticehurst, R. H. Purcell, A. J. Buckler-White, and B. M. Baroudy. 1987. Complete nucleotide sequence of wild-type hepatitis A virus: comparison with different strains of hepatitis A virus and other picornaviruses. *J. Virol.* **61**:50–59.
8. Committee on Enteroviruses. 1957. The enteroviruses. *Am. J. Public Health* **47**:1556–1566.
9. Committee on the Enteroviruses. 1962. Classification of human enteroviruses. *Virology* **16**:501–504.
10. Contreras, G., V. H. Barnett, and J. L. Melnick. 1952. Identification of coxsackie viruses by immunological methods and their classification into 16 antigenically distinct types. *J. Immunol.* **69**:395–414.
11. Dahllund, L., L. Nissinen, T. Pulli, V.-P. Hyttinen, G. Stanway, and T. Hyypiä. 1995. The genome of echovirus 11. *Virus Res.* **35**:215–222.
12. De, L., B. Nottay, C.-F. Yang, B. P. Holloway, M. Pallansch, and O. Kew. 1995. Identification of vaccine-related polioviruses by hybridization with specific RNA probes. *J. Clin. Microbiol.* **33**:562–571.
13. De, L., C.-F. Yang, E. da Silva, J. Boshell, P. Cáceres, J. Ruiz Gómez, M. Pallansch, and O. Kew. 1997. Genotype-specific RNA probes for direct identification of wild polioviruses by blot hybridization. *J. Clin. Microbiol.* **35**:2834–2840.
14. Dorner, A. J., L. F. Dorner, G. R. Larsen, E. Wimmer, and C. W. Anderson. 1982. Identification of the initiation site of poliovirus polypeptide synthesis. *J. Virol.* **42**:1017–1028.
15. Earle, J. A., R. A. Skuce, C. S. Fleming, E. M. Hoey, and S. J. Martin. 1988. The complete nucleotide sequence of a bovine enterovirus. *J. Gen. Virol.* **69**:253–263.
16. Emini, E. A., M. Elzinga, and E. Wimmer. 1982. Carboxy-terminal analysis of poliovirus proteins: termination of poliovirus RNA translation and location of unique poliovirus polypeptide cleavage sites. *J. Virol.* **42**:194–199.
17. Felsenstein, J. 1993. PHYLIP: phylogeny inference package, version 3.5c (computer program). University of Washington, Seattle.
18. Grandien, M., M. Forsgren, and A. Ehrnst. 1989. Enteroviruses and reoviruses, p. 513–569. In E. H. Lennette and N. J. Schmidt (ed.), *Diagnostic procedures for viral, rickettsial, and chlamydial infections*, 6th ed. American Public Health Association, Washington, D.C.
19. Gratsch, T. E., and V. F. Righthand. 1994. Construction of a recombinant cDNA of echovirus 6 that established a persistent in vitro infection. *Virology* **201**:341–348.
20. Harris, L. F., R. E. Haynes, H. G. Cramblett, R. M. Conant, and G. R. Jenkins. 1973. Antigenic analysis of echoviruses 1 and 8. *J. Infect. Dis.* **127**:63–68.
21. Hogle, J. M., M. Chow, and D. J. Filman. 1985. Three-dimensional structure of poliovirus at 2.9 Å resolution. *Science* **229**:1358–1365.
22. Hughes, P. J., C. North, P. D. Minor, and G. Stanway. 1989. The complete nucleotide sequence of coxsackievirus A21. *J. Gen. Virol.* **70**:2943–2952.
23. Huttunen, P., J. Santti, T. Pulli, and T. Hyypiä. 1996. The major echovirus group is genetically coherent and related to coxsackie B viruses. *J. Gen. Virol.* **77**:715–725.
24. Hyypiä, T., C. Horsnell, M. Maaronen, M. Khan, N. Kalkkinen, P. Auvinen, L. Kinnunen, and G. Stanway. 1992. A distinct picornavirus group identified by sequence analysis. *Proc. Natl. Acad. Sci. USA* **89**:8847–8851.
25. Hyypiä, T., T. Hovi, N. J. Knowles, and G. Stanway. 1997. Classification of enteroviruses based on molecular and biological properties. *J. Gen. Virol.* **78**:1–11.
26. Iizuka, N., S. Kuge, and S. Nomoto. 1987. Complete nucleotide sequence of the genome of coxsackievirus B1. *Virology* **156**:64–73.
27. Inoue, T., T. Suzuki, and K. Sekiguchi. 1989. The complete nucleotide sequence of swine vesicular disease virus. *J. Gen. Virol.* **70**:919–934.
28. Jenkins, O., J. D. Booth, P. D. Minor, and J. W. Almond. 1987. The complete nucleotide sequence of coxsackievirus B4 and its comparison to other members of the Picornaviridae. *J. Gen. Virol.* **68**:1835–1848.
29. Kapsenberg, J. G. 1988. *Picornaviridae: the enteroviruses (polioviruses, coxsackieviruses, echoviruses)*, p. 692–722. In E. H. Lennette, P. Halonen, and F. A. Murphy (ed.), *Laboratory diagnosis of infectious diseases. Principles and practice*, vol. II. Viral, rickettsial, and chlamydial diseases. Springer-Verlag, New York, N.Y.
30. Kew, O. M., M. N. Mulders, G. Y. Lipskaya, E. E. da Silva, and M. A. Pallansch. 1995. Molecular epidemiology of polioviruses. *Semin. Virol.* **6**:401–414.
31. Kilpatrick, D. R., B. Nottay, C.-F. Yang, S.-J. Yang, M. N. Mulders, B. P. Holloway, M. Pallansch, and O. M. Kew. 1996. Group-specific identification of polioviruses by PCR using primers containing mixed-base or deoxyninosine residues at positions of codon degeneracy. *J. Clin. Microbiol.* **34**:2990–2996.
32. Kilpatrick, D. R., B. Nottay, C.-F. Yang, S.-J. Yang, E. da Silva, S. Peñaranda, M. Pallansch, and O. M. Kew. 1998. Serotype-specific identification of polioviruses by PCR using primers containing mixed-base or deoxyninosine residues at positions of codon degeneracy. *J. Clin. Microbiol.* **36**:352–357.
33. Kishino, H., and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in *Hominoidia*. *J. Mol. Evol.* **29**:170–179.
34. Kitamura, N., B. L. Semler, P. G. Rothberg, G. R. Larsen, C. J. Adler, A. J. Dorner, E. A. Emini, R. Hanecak, J. J. Lee, S. Van der Werf, C. W. Anderson, and E. Wimmer. 1981. Primary structure, gene organization and polypeptide expression of poliovirus RNA. *Nature* **291**:547–553.
35. Kopecka, H., B. Brown, and M. A. Pallansch. 1995. Genotypic variation in coxsackievirus B5 isolates from three different outbreaks in the United States. *Virus Res.* **38**:125–136.
36. Kraus, W., H. Zimmermann, A. Zimmermann, H. J. Eggers, and B. Nelsen-Salz. 1995. Infectious cDNA clones of echovirus 12 and a variant resistant against the uncoating inhibitor rhodanine differ in seven amino acids. *J. Virol.* **69**:5853–5858.
37. La Monica, N., C. Meriam, and V. R. Racaniello. 1986. Mapping of sequences required for mouse neurovirulence of poliovirus type 2 Lansing. *J. Virol.* **57**:515–525.
38. Lindberg, A. M., P. O. Ståhlhandske, and U. Pettersson. 1987. Genome of coxsackievirus B3. *Virology* **156**:50–63.
39. Luo, M., G. Vriend, G. Kamer, I. Minor, E. Arnold, M. G. Rossman, U. Boege, D. G. Scraba, G. M. Duke, and A. C. Palmenberg. 1987. The atomic structure of mengovirus at 3.0 Å resolution. *Science* **235**:182–191.
40. Mateu, M. G. 1995. Antibody recognition of picornaviruses and escape from neutralization: a structural view. *Virus Res.* **38**:1–24.
41. McNally, R. M., J. A. Earle, M. McIlhatton, E. M. Hoey, and S. J. Martin.

1994. The nucleotide sequence of the 5' non-coding and capsid coding genome regions of two bovine enterovirus strains. *Arch. Virol.* **139**:287–299.
42. Melnick, J. L., I. Tagaya, and H. von Magnus. 1974. Enteroviruses 69, 70, and 71. *Intervirology* **4**:369–370.
43. Melnick, J. L. 1996. Enteroviruses: polioviruses, coxsackieviruses, echoviruses, and newer enteroviruses, p. 655–712. *In* B. N. Fields, D. M. Knipe, P. M. Howley, R. M. Chanock, J. L. Melnick, T. P. Monath, B. Roizman, and S. E. Straus (ed.), *Fields virology*, 3rd ed. Lippincott-Raven Publishers, Philadelphia, Pa.
44. Minor, P. D. 1990. Antigenic structure of picornaviruses. *Curr. Top. Microbiol. Immunol.* **161**:121–154.
45. Minor, P. D., F. Brown, E. Domingo, E. Hoey, A. King, N. Knowles, S. Lemon, A. Palmenberg, R. R. Rueckert, G. Stanway, E. Wimmer, and M. Yin-Murphy. 1995. Family Picornaviridae, p. 329–336. *In* F. A. Murphy, C. M. Fauquet, D. H. L. Bishop, S. A. Ghabrial, A. W. Jarvis, G. P. Martelli, M. A. Mayo, and M. D. Summers (ed.), *Virus taxonomy. Classification and nomenclature of viruses*. Springer-Verlag, Vienna, Austria.
46. Morens, D. M., and M. A. Pallansch. 1995. Epidemiology, p. 3–23. *In* H. A. Rotbart (ed.), *Human enterovirus infections*. ASM Press, Washington, D.C.
47. Nomoto, A., T. Omata, H. Toyoda, S. Kuge, H. Horie, Y. Kataoka, Y. Genba, Y. Nakano, and N. Imura. 1982. Complete nucleotide sequence of the attenuated poliovirus Sabin 1 strain genome. *Proc. Natl. Acad. Sci. USA* **79**:5793–5797.
48. Oberste, M. S., S. C. Weaver, D. M. Watts, and J. F. Smith. 1998. Identification and genetic analysis of Panama-genotype Venezuelan equine encephalomyelitis virus subtype ID in Peru. *Am. J. Trop. Med. Hyg.* **58**:41–46.
49. Oberste, M. S., K. Maher, and M. A. Pallansch. 1998. Molecular phylogeny of all human enterovirus serotypes based on comparison of sequences at the 5' end of the region encoding VP2. *Virus Res.* **58**:35–43.
50. Oberste, M. S., K. Maher, and M. A. Pallansch. 1998. Complete sequence of echovirus 23 and its relationship to echovirus 22 and other human enteroviruses. *Virus Res.* **56**:217–223.
51. Peng, J. H., J. W. McCauley, R. P. Kitching, and N. J. Knowles. Unpublished data.
52. Pöyry, T., T. Hyypiä, C. Horsnell, L. Kinnunen, T. Hovi, and G. Stanway. 1994. Molecular analysis of coxsackievirus A16 reveals a new genetic group of enteroviruses. *Virology* **202**:982–987.
53. Pöyry, T., L. Kinnunen, T. Hyypiä, B. Brown, C. Horsnell, T. Hovi, and G. Stanway. 1996. Genetic and phylogenetic clustering of enteroviruses. *J. Gen. Virol.* **77**:1699–1717.
54. Pulli, T., P. Koskimies, and T. Hyypiä. 1995. Molecular comparison of coxsackie A virus serotypes. *Virology* **211**:30–38.
55. Racaniello, V. R., and D. Baltimore. 1981. Molecular cloning of poliovirus cDNA and determination of the complete nucleotide sequence of the viral genome. *Proc. Natl. Acad. Sci. USA* **78**:4887–4891.
56. Ramos-Alvarez, M., and A. B. Sabin. 1954. Characteristics of poliomyelitis and other enteric viruses recovered in tissue culture from healthy American children. *Proc. Soc. Exp. Biol. Med.* **87**:655–661.
57. Reimann, B.-Y., R. Zell, and R. Kandolf. 1991. Mapping of a neutralizing antigenic site of coxsackievirus B4 by construction of an antigen chimera. *J. Virol.* **65**:3475–3480.
58. Rossman, M. G., A. Arnold, J. W. Erickson, E. A. Frankengerger, J. P. Griffith, H.-J. Hecht, J. E. Johnson, G. Kamer, M. Luo, A. G. Mosser, R. R. Rueckert, B. Sherry, and G. Vriend. 1985. Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* **317**:145–153.
59. Ryan, M. D., O. Jenkins, P. J. Hughes, A. Brown, N. J. Knowles, D. Booth, P. D. Minor, and J. W. Almond. 1990. The complete nucleotide sequence of enterovirus type 70: relationships with other members of the Picornaviridae. *J. Gen. Virol.* **71**:2291–2299.
60. Schmidt, N. J., and E. H. Lennette. 1970. Antigenic variants of coxsackievirus A24. *Am. J. Epidemiol.* **91**:99–109.
61. Skern, T., W. Sommergruber, D. Blaas, P. Gruendler, F. Fraundorfer, C. Pieler, I. Fogy, and E. Kuechler. 1985. Human rhinovirus 2: complete sequence and proteolytic processing signals in the capsid protein region. *Nucleic Acids Res.* **13**:2111–2126.
62. Stanway, G., P. J. Hughes, R. C. Mountford, P. Reeve, P. D. Minor, G. C. Schild, and J. W. Almond. 1984. Comparison of the complete nucleotide sequences of the genomes of the neurovirulent poliovirus P3/Leon/37 and its attenuated Sabin vaccine derivative P3/Leon/12a1b. *Proc. Natl. Acad. Sci. USA* **81**:1539–1543.
63. Stanway, G., P. J. Hughes, R. C. Mountford, P. D. Minor, and J. W. Almond. 1984. The complete nucleotide sequence of a common cold virus: human rhinovirus 14. *Nucleic Acids Res.* **12**:7859–7875.
64. Strimmer, K., and A. von Haeseler. 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**:964–969.
65. Supanaranond, K., N. Takeda, and S. Yamazaki. 1992. The complete nucleotide sequence of a variant of coxsackievirus A24, an agent causing acute hemorrhagic conjunctivitis. *Virus Genes* **6**:149–158.
66. Toyoda, H., M. Kohara, Y. Kataoka, T. Suganuma, T. Omata, N. Imura, and A. Nomoto. 1984. Complete nucleotide sequences of all three poliovirus serotype genomes. Implication for genetic relationship, gene function and antigenic determinants. *J. Mol. Biol.* **174**:561–585.
67. Van Regenmortel, M. H. V., D. H. L. Bishop, C. M. Fauquet, M. A. Mayo, J. Maniloff, and C. H. Calisher. 1997. Guidelines to the demarcation of virus species. *Arch. Virol.* **142**:1505–1518.
68. Wenner, H. A., P. Harmon, A. M. Behbehani, H. Rouhandeh, and P. Kamitsuka. 1967. The antigenic heterogeneity of type 30 echoviruses. *Am. J. Epidemiol.* **85**:240–249.
69. Yang, C.-F., L. De, B. P. Holloway, M. A. Pallansch, and O. M. Kew. 1991. Detection and identification of vaccine-related polioviruses by the polymerase chain reaction. *Virus Res.* **20**:159–179.
70. Yang, C.-F., L. De, S.-J. Yang, J. Ruiz Gómez, J. Ramiro Cruz, B. P. Holloway, M. A. Pallansch, and O. M. Kew. 1992. Genotype-specific in vitro amplification of sequences of the wild type 3 polioviruses from Mexico and Guatemala. *Virus Res.* **24**:277–296.
71. Yin-Murphy, M. 1972. The picornaviruses of acute haemorrhagic conjunctivitis: a comparative study. *Southeast Asian J. Trop. Med. Public Health* **3**:303–309.
72. Yohn, D. S., and W. M. Hammon. 1960. ECHO-4 viruses: improved methods and strain selection for identification and serodiagnosis. *Proc. Soc. Exp. Biol. Med.* **105**:55–60.
73. Zell, R., and A. Stelzner. 1997. Application of genome sequence information to the classification of bovine enteroviruses: the importance of 5'- and 3'-nontranslated regions. *Virus Res.* **51**:213–229.
74. Zhang, G., G. Wilsden, N. J. Knowles, and J. W. McCauley. 1993. Complete nucleotide sequence of a coxsackie B5 virus and its relationship to swine vesicular disease virus. *J. Gen. Virol.* **74**:845–853.
75. Zimmermann, H., H. J. Eggers, A. Zimmermann, W. Kraus, and B. Nelsen-Salz. 1995. Complete nucleotide sequence and biological properties of an infectious clone of prototype echovirus 9. *Virus Res.* **39**:311–319.
76. Zimmermann, H., H. Eggers, and B. Nelsen-Salz. 1996. Molecular cloning and sequence determination of the complete genome of the virulent echovirus 9 strain Barty. *Virus Genes* **12**:149–154.